

Diseños de investigación y Análisis de Datos. Septiembre 2011, soluciones al modelo A.

A continuación se presentan 3 situaciones. Cada situación viene seguida por una serie de preguntas referidas a la misma y de preguntas teóricas generales.

**SITUACIÓN 1:** “La formación alcanzada por la población adulta española ha mejorado, de forma continua, en los últimos 10 años. Desde 1998 el porcentaje de españoles de 25 a 64 años que poseen estudios superiores a los obligatorios ha pasado de 33% al 51% en 2008. En la misma proporción ha disminuido, por tanto, el porcentaje de españoles que sólo poseen estudios obligatorios, que ha pasado del 67% en 1998 al 49% en 2008” (Panorama de la Educación. Informe OCDE 2010). Suponga que usted quiere estudiar si estos datos son los que realmente existen en su Comunidad, para lo que utiliza una muestra aleatoria de 900 adultos con edades comprendidas entre 25 y 64 años, encontrando que 378 de ellos tienen solo los estudios obligatorios.

Los datos de los que partimos son:

|      | Estudios superiores a los obligatorios | Estudios obligatorios | Muestra                                |
|------|--|-----------------------|--|
| 1998 | 33%                                    | 67%                   | $n = 900$                              |
| 2008 | 51%                                    | 49%                   | 378 personas con estudios obligatorios |

Proporción de personas en la muestra con estudios obligatorios:  $p = \frac{378}{900} = 0,42$

1.- Si decide trabajar con un nivel de confianza del 95% y se establece el error máximo de estimación en el 3%, ¿cuál debe ser el tamaño de la muestra?: **A) 1040**; B) 544; C) 789.

El error máximo viene dado por la expresión:

$$E = Z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}$$

Despejando “n” de la ecuación anterior, obtenemos:

$$n = p(1-p) \frac{Z_{\alpha/2}^2}{E^2} = 0,42 \times 0,58 \times \frac{1,96^2}{0,03^2} = 1039,79 \approx 1040$$

2.- El tamaño de la muestra que se requiere para estimar la proporción poblacional con un nivel de confianza previamente fijado: A) aumenta cuando aumenta el error en la estimación; **B) aumenta cuando disminuye el error en la estimación**; C) no depende del error en la estimación.

Simplemente observando la fórmula con la que obtenemos el tamaño de la muestra concluimos que la respuesta correcta es “B”.

3.- Sean “X” e “Y” dos estimadores de un mismo parámetro, si la distribución de probabilidad de “X” tiene menor variabilidad que “Y”: A) el estimador “Y” es más eficiente; **B) el estimador “X” es más eficiente**; C) el estimador “X” es más suficiente.

*Pregunta teórica. Página 33 del libro de texto*

4.- Trabajando con un nivel de confianza del 95%, el intervalo de confianza de la proporción de adultos de su Comunidad que solo tienen estudios obligatorios, se encuentra entre: A) 0,378 y 0,462; **B) 0,388 y 0,452**; C) 0,376 y 0,461.

$$p \pm Z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}} \rightarrow 0,42 \pm 1,96 \sqrt{\frac{0,42 \times 0,58}{900}} \rightarrow \begin{cases} 0,388 \\ 0,452 \end{cases}$$

5.- El nivel de confianza se refiere a: **A) La probabilidad de que el parámetro se encuentre dentro de un intervalo de confianza**; B) El valor inverso del error típico; C) La probabilidad de que el estadístico se encuentre dentro de la distribución muestral del parámetro.

*Pregunta teórica. Tema 1 del libro de texto.*

6.- Si el investigador quiere confirmar que la proporción de adultos de su Comunidad que solo tienen los estudios obligatorios es significativamente inferior que el indicador de España en 2008, la hipótesis alternativa es: **A)  $H_1: \pi < 0,49$** ; B)  $H_1: \pi > 0,49$ ; C)  $H_1: \pi \leq 0,42$ .

*La respuesta correcta es “A”, puesto que las hipótesis que debemos plantear son:*

$$H_0: \pi \geq 0,49$$

$$H_1: \pi < 0,49$$

7.- Si el investigador cree que la proporción de adultos que solo tienen los estudios obligatorios es inferior que el indicador de España en 2008, el estadístico de contraste y el nivel p-crítico para analizar si la diferencia es estadísticamente significativa es: A) 3,34 ( $p=0,0008$ ); **B) -4,2 ( $p<0,0002$ )**; C) -2,33 ( $p=0,0198$ );

$$Z = \frac{0,42 - 0,49}{\sqrt{\frac{0,49 \times 0,51}{900}}} = -4,20 \xrightarrow{\text{Tabla curva normal}} p < 0,0002$$

8.- Respecto a la hipótesis del investigador analizada en las preguntas anteriores, el investigador concluye que la proporción de adultos que solo tienen estudios obligatorios en su Comunidad, es: **A) inferior a la del resto de España con un nivel de significación de 0,01**; B) la misma que la del resto de España con un nivel de significación de 0,05; C) superior a la del resto de España con un nivel de significación de 0,01.

*Obviamente, la respuesta correcta es “A”.*

9.- A la probabilidad de rechazar la hipótesis nula cuando es falsa se le denomina: A) nivel de confianza; **B) potencia del contraste**; C) nivel p-crítico.

*Pregunta teórica. Página 61 del libro de texto.*

**SITUACIÓN 2:** En una tarea sobre decisión léxica con estímulos visuales en la que se miden tiempos de reacción, un investigador dispone de una muestra aleatoria de 26 sujetos. Utiliza dos grupos de palabras y parte de la hipótesis de que los tiempos de reacción serán más cortos ante palabras de alta frecuencia léxica (aquellas que utilizamos más

frecuentemente en el lenguaje, Grupo 1) que ante palabras de baja frecuencia léxica (Grupo 2). Tras realizar el experimento observa que el tiempo de reacción medio para el Grupo 1 fue igual a 612,04 ms, mientras que para el Grupo 2 fue igual a 645 ms, siendo la cuasivarianza de las diferencias igual a 6656 ms<sup>2</sup>. Asumiendo que se cumplen los supuestos para realizar un contraste paramétrico, conteste a las siguientes cuestiones:

10.- Para comprobar si la diferencia entre dos medias es significativa, calculamos: **A) el estadístico de contraste**; B) el tamaño del efecto; C) las dos opciones anteriores son verdaderas.

*Pregunta teórica. Página 156 del libro de texto.*

11.- El nivel de medida de la variable dependiente es: A) ordinal; B) de intervalo; **C) de razón**.

*La variable dependiente es el tiempo de reacción, cuyo nivel de medida es de razón.*

12.- La hipótesis nula es: **A)  $H_0: \mu_1 - \mu_2 \geq 0$** ; B)  $H_0: \mu_1 - \mu_2 < 0$ ; C)  $H_0: \mu_1 - \mu_2 \leq 0$ .

*La respuesta correcta es "A", debemos de plantear un contraste de hipótesis sobre dos medias para dos muestras relacionadas, siendo las hipótesis:*

$$\begin{aligned} H_0: \mu_1 - \mu_2 &\geq 0 \\ H_1: \mu_1 - \mu_2 &< 0 \end{aligned}$$

13.- La hipótesis alternativa es: A)  $H_1: \mu_1 - \mu_2 \geq 0$ ; **B)  $H_1: \mu_1 - \mu_2 < 0$** ; C)  $H_1: \mu_1 - \mu_2 > 0$ .

14.- El valor absoluto del estadístico de contraste es: A) 1,708; **B) 2,06**; C) 2,485.

*La respuesta correcta es "B". El estadístico de contraste es:*

$$T = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{S_d^2}{n}}} = \frac{612,04 - 645}{\sqrt{\frac{6656}{26}}} = -2,06$$

15.- Para un nivel de confianza del 95%, el valor absoluto del valor crítico es: **A) 1,708**; B) 2,06; C) 2,485.

*Acudiendo a la tabla t de Student, observamos que la respuesta correcta es "A".*

16.- El valor del nivel p-crítico es: A) 0,05; **B) 0,025**; C) 0,01.

*Acudiendo a la tabla t de Student, observamos que la respuesta correcta es "B".*

17.- El resultado del experimento: A) no es significativo; **B) es significativo al nivel de confianza del 95%**; C) es significativo al 95% y al 99%.

*La respuesta correcta es "B", puesto que los resultados son significativos para el nivel de confianza del 95% ( $-2,06 < -1,708$ ) pero no son significativos al nivel de confianza del 99% ( $-2,06 > -2,485$ ).*

**SITUACIÓN 3.** Un investigador sospecha que los tiempos de coagulación de la sangre difieren en función del tipo de dieta. Para ello, ha realizado un experimento donde ha formado varios grupos de ratones, sometiendo a cada grupo a distintos tipos de dietas. Tras analizar los datos mediante un ANOVA, parte de los siguientes resultados:

| Fuentes de variación | Sumas de cuadrados | Grados de libertad | Medias cuadráticas | F |
|----------------------|--------------------|--------------------|--------------------|---|
| Entre niveles        | 228                | 3                  |                    |   |
| Dentro de niveles    | 112                | 20                 |                    |   |
| Total                | 340                | 23                 |                    |   |

Conociendo los grados de libertad, deducimos el número de niveles, número de sujetos en cada nivel y número total de sujetos:

$$I - 1 = 3 \rightarrow I = 4; \quad N - 1 = 23 \rightarrow N = 24; \quad n = \frac{N}{I} = \frac{24}{4} = 6$$

Completamos la tabla de ANOVA, que queda:

| F.V.         | S.C. | g.l. | M.C.           | F                |
|--------------|------|------|----------------|------------------|
| Entre grupos | 228  | 3    | $228/3 = 76$   | $76/5,6 = 13,57$ |
| Error        | 112  | 20   | $112/20 = 5,6$ |                  |
| Total        | 340  | 23   |                |                  |

18.- Asumiendo un diseño equilibrado, ¿cuántos animales utilizó por grupo? **A) 6**; B) 4; C) 5.

La respuesta correcta es "A".

19.- ¿Cuántos niveles tiene el factor "Tipo de dieta"? A) 6; **B) 4**; C) 3.

La respuesta correcta es "B".

20.- La hipótesis nula planteada en este diseño es: **A)  $H_0: \mu_i = \mu_j$  para todo  $i \neq j$** ; B)  $H_0: \mu_i \neq \mu_j$  para todo  $i = j$ ; C)  $Y_{ij} = \mu + \alpha_{ij} + \varepsilon_{ij}$ .

Pregunta teórica. Página 208 del libro de texto.

21.- El valor de la F para el factor "Tipo de dieta" fue: A) 8,660; B) 2,380; **C) 13,57**.

La respuesta correcta es "C".

22.- Siendo  $\alpha = 0.05$ , el valor de la F crítica para el factor "Tipo de dieta" fue: **A) 3,098**; B) 2,380; C) 4,938.

Para un nivel de confianza del 95%, mirando en las tablas F de Fisher para 3 y 20 grados de libertad, observamos que la F crítica es: 3,098, luego la alternativa correcta es "A".

23.- ¿Tendremos que realizar comparaciones *a posteriori*? **A) Sí**; B) No porque tenemos más de 2 niveles del factor; C) No porque las comparaciones *a posteriori* solo se aplican en diseños intrasujetos.

Tendremos que hacer comparaciones "a posteriori" porque hemos rechazado la hipótesis nula que planteaba la igualdad entre las medias ( $13,57 > 3,098 \rightarrow H_1$ ).

24.- Con un nivel de confianza del 95%, ¿Cuál será el valor mínimo de la diferencia entre dos medias cualesquiera en este experimento para que resulten significativas según Tukey ( $\alpha = 0,05$ )? A) 12,39; B) 0,59; **C) 3,83**.

$$HSD_{Tukey} = q_{\alpha, gl\ error, k} \sqrt{\frac{MC_{error}}{n}} = 3,96 \sqrt{\frac{5,6}{6}} = 3,83$$

Luego la respuesta correcta es "C".

25.- Cuando se aplica un Análisis de la Varianza de un factor, las varianzas de los grupos a comparar deben de cumplir el supuesto de: A) complementariedad y exhaustividad; **B) homocedasticidad**; C) independencia.

Pregunta teórica. Página 232.